

Selection and Drift in Subdivided Populations: A Straightforward Method for Deriving Diffusion Approximations and Applications Involving Dominance, Selfing and Local Extinctions

Denis Roze^{*,†,1} and François Rousset^{*}

^{*}*Institut des Sciences de l'Evolution, Université Montpellier II, 34095 Montpellier, France and* [†]*Centre d'Etudes sur le Polymorphisme des Microorganismes, Institut de Recherche pour le Développement, 34394 Montpellier, France*

Manuscript received February 11, 2003

Accepted for publication May 19, 2003

ABSTRACT

Population structure affects the relative influence of selection and drift on the change in allele frequencies. Several models have been proposed recently, using diffusion approximations to calculate fixation probabilities, fixation times, and equilibrium properties of subdivided populations. We propose here a simple method to construct diffusion approximations in structured populations; it relies on general expressions for the expectation and variance in allele frequency change over one generation, in terms of partial derivatives of a "fitness function" and probabilities of genetic identity evaluated in a neutral model. In the limit of a very large number of demes, these probabilities can be expressed as functions of average allele frequencies in the metapopulation, provided that coalescence occurs on two different timescales, which is the case in the island model. We then use the method to derive expressions for the probability of fixation of new mutations, as a function of their dominance coefficient, the rate of partial selfing, and the rate of deme extinction. We obtain more precise approximations than those derived by recent work, in particular (but not only) when deme sizes are small. Comparisons with simulations show that the method gives good results as long as migration is stronger than selection.

THE geographical distribution of a population has many consequences on its evolution. The nature of the habitat (fragmented or continuous) and the dispersal patterns of individuals across space affect the relative importance of the different processes changing allele frequencies (selection, genetic drift...) and therefore have an effect on several genetical features of populations, such as the amount of genetic diversity and the strength of inbreeding depression. Spatial structure may also affect the timescale of evolution; indeed, the rate of adaptation of a population depends critically on the probability that new advantageous mutations will go to fixation and on the time length of the fixation process.

Different models have been used to calculate the probability of fixation of an advantageous mutation in a spatially structured population, dealing with different types of population structure, and different modes of selection. For some simple models, MARUYAMA (1970) found that in a population subdivided into demes of constant size that never go extinct, spatial structure has no influence on fixation probabilities: a new mutation has the same probability of fixation as if the population was not fragmented. This conclusion has been supported by more recent work (CHERRY and WAKELEY 2003).

SLATKIN (1981) used a low-migration limit to investigate the joint effects of dominance and spatial structure on fixation probabilities of advantageous mutations. In his model, the fixation of an allele within a deme occurs much faster than its transmission between demes, so that fixation is assumed to take place independently in each deme (see also TAKAHATA 1991). He showed that the probability of fixation of recessive mutations is increased in this low-migration limit, compared to the case of a panmictic population; indeed, population structure increases the probability that an advantageous homozygous genotype is created before the mutation is lost. However, the probability of fixation of dominant mutations is decreased. BARTON (1993) worked on another model to study the effect of catastrophic extinctions of demes and showed that extinctions can substantially reduce the fixation probability of beneficial alleles. More recently, several authors used diffusion approximations to derive fixation probabilities and fixation times in metapopulations. CHERRY (2003a,b), CHERRY and WAKELEY (2003), and WAKELEY (2003) have shown that diffusion methods can be used in island models of population structure and studied the joint effects of dominance, population structure, and local extinctions on fixation probabilities and times of mutant alleles. WHITLOCK (2003) has presented another approach, which he also applies to the stepping-stone model of population structure.

In this article we provide a different method to construct diffusion approximations, which is in several ways

¹*Corresponding author:* Centre d'Etudes sur le Polymorphisme des Microorganismes, UMR CNRS-IRD 9926, Institut de Recherche pour le Développement, 911 avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France. E-mail: roze@mpl.ird.fr

more general than previous ones. The method consists of expressing the mean of allele frequency change over one generation in terms of partial derivatives of a “fitness function” and probabilities of genetic identity computed in a neutral model. We make no assumption on the particular form of the fitness function; indeed, the method can be applied to many different selection models (including multilevel selection). We then take the example, considered by previous authors, of selection through effects on fecundity (different genotypes producing different numbers of gametes) and consider the effects of dominance, self-fertilization, and local extinctions on the probability of fixation of new mutations. We confirm some previous approximations; however, our method gives more precise results when migration is strong, when deme size is small, and also when extinctions occur. As discussed later, the diffusion method gives satisfying results as long as migration rates are not too small; with dominance, Slatkin’s method performs better when migration is very low.

LIFE CYCLE

Throughout this article, we use an island model of population structure, with possible local extinctions. We assume that the population is subdivided into n demes. At the beginning of the life cycle, each deme contains N adult individuals and has a probability e of becoming extinct. In the surviving demes, individuals produce a very large number of gametes and die; in practice we treat the number of gametes produced as if it were infinite. Gamete fusion may precede or follow migration (zygotic or gametic migration); we first do not make any assumption about the breeding system, but later we consider the cases of random union of gametes and of partial self-fertilization at a rate α . Migrants produced by a deme can reach any other deme with the same probability. The migration rate m is defined as the proportion of migrant zygotes (or gametes) present in any surviving deme once migration has occurred (“backward” migration rate). We assume that migrants arriving in extinct demes do not survive, each extinct deme being recolonized later by a finite number k of individuals (SLATKIN 1977). These recolonizers come from “propagules” that formed in each surviving deme after the first round of dispersal. Again, recolonization can be zygotic (if it occurs after fertilization) or gametic (if it occurs before; in this case each extinct deme is recolonized by $2k$ gametes). We also define a parameter ϕ measuring the probability that two colonizers come from the same deme (this is specified later). We assume that colonizers reproduce immediately, producing a very large number of zygotes. Finally, in all demes N individuals are sampled randomly among all the zygotes to form the new adult generation; exactly N adults are always in each deme, recolonized or not. Thus, recolonizers experience two rounds of migration and reproduction within

one “generation.” This way of representing the dynamics of recolonization is obviously simplistic, but may be seen as a first step in understanding the consequences of extinction and recolonization (SLATKIN 1977; WADE and McCauley 1988; WHITLOCK and BARTON 1997; PANNELL and CHARLESWORTH 1999); more complex models may be tractable within the framework presented here (ROUSSET 2004, Chaps. 10 and 11).

THE DIFFUSION APPROXIMATION

Diffusion theory has provided good approximations for probabilities of fixation of mutant alleles (KIMURA 1964). In this section we briefly review the methodology in the case of a panmictic population containing two alleles a and A . We call p the frequency of A and δp the change in frequency of A over a short time interval δt . We first suppose that time and frequency are continuous and define

$$M_{\delta p} = E[\delta p | p] \quad (1)$$

$$V_{\delta p} = E[(\delta p)^2 | p], \quad (2)$$

where $E[x]$ is the expectation of x . To use the diffusion method we need the following limits:

$$\mathcal{M}_{\delta p} = \lim_{\delta t \rightarrow 0} (M_{\delta p} / \delta t) \quad (3)$$

$$\mathcal{V}_{\delta p} = \lim_{\delta t \rightarrow 0} (V_{\delta p} / \delta t). \quad (4)$$

If these limits are finite, and if the higher-order moments of δp are negligible, the probability of fixation of A when present in frequency p , $u(p)$, is given by the differential equation

$$\mathcal{M}_{\delta p} \frac{du}{dp} + \frac{\mathcal{V}_{\delta p}}{2} \frac{d^2 u}{dp^2} = 0 \quad (5)$$

(KARLIN and TAYLOR 1981) with solution

$$u(p) = \frac{\int_0^p G(x) dx}{\int_0^1 G(x) dx}, \quad (6)$$

where

$$G(x) = \exp \left[- \int \frac{2\mathcal{M}_{\delta x}}{\mathcal{V}_{\delta x}} dx \right]. \quad (7)$$

To apply this continuous model to the case of a population with discrete generations, one states that time is measured in units of N generations, where N is the population size, and that the short time interval δt corresponds to one generation. Then $\delta t = 1/N$, and $\mathcal{M}_{\delta p}$ and $\mathcal{V}_{\delta p}$ are the limits as N tends to infinity of $N \times M_{\delta p}$ and $N \times V_{\delta p}$, respectively (KARLIN and TAYLOR 1981).

In the simple case of genic selection where A has a selective advantage s over a in an ideal haploid Wright-Fisher population of size N , we have

$$M_{\delta p} = spq + o(s) \quad (8)$$

$$V_{\delta p} = \frac{pq}{N} + o(1/N) + o(s), \quad (9)$$

with $q = 1 - p$ (EWENS 1979; KARLIN and TAYLOR 1981). For $M_{\delta p}$ and $V_{\delta x}$ to be finite, we need to assume that s tends to zero as N tends to infinity and that the product Ns has a finite limit, say σ . Then Equations 6 and 7 are valid, with $M_{\delta p} = \sigma pq$ and $V_{\delta x} = pq$. These expressions of $M_{\delta p}$ and $V_{\delta x}$ lead to the classical result (e.g., EWENS 1979)

$$u(p) = \frac{1 - \exp[-2Ns p]}{1 - \exp[-2Ns]}, \quad (10)$$

where N is assumed to be large and s small. Despite all the approximations done, this formula proves very accurate for values of s as high as 0.1, even when N is as small as 10 (e.g., CARR and NASSAR 1970; GALE 1990).

In the following, we consider an island model of population structure and look for an expression for the probability of fixation of a mutant allele A in a population initially fixed for allele a ; p is now the frequency of A in the whole population, while the vector $\mathbf{p} = (p_1, \dots, p_n)$ gives the frequencies of A in different demes. We suppose that the number of demes (n) is large and define $M_{\delta p}$ and $V_{\delta p}$ as the limits as n tends to infinity of the products $n \times M_{\delta p}$ and $n \times V_{\delta p}$, where $M_{\delta p}$ and $V_{\delta p}$ are still the first two moments of the change of A over one generation. Using a simple argument, we show that these limits can be expressed as functions of the average frequency of A in the population (p). A rigorous demonstration (including the proof that higher-order moments vanish in the limit as n tends to infinity) would follow the general argument of ETHIER and NAGYLAKI (1980) for Markov chains with two timescales. WAKELEY (2003) explains how their equations apply to island models of structured populations. We do not repeat such arguments here, but we offer an alternative presentation of the same general idea.

EXPECTED CHANGE IN ALLELE FREQUENCY

We use the direct fitness method (TAYLOR and FRANK 1996; ROUSSET and BILLIARD 2000) to express $M_{\delta p}$ as a function of probabilities of identity and partial derivatives of a fitness function. The parameters and variables of the model are summarized in Tables 1 and 2. In all the following, the subscript ij refers to the j th adult individual of the i th deme. The fitness W_{ij} of this individual is defined as the expected number of its gametes that will participate to the next adult generation. Fitness may depend on the phenotype of the individual, on the mean phenotype of the individuals present in its deme, and on the mean phenotype in the whole population. We call z_{aa} the phenotype of homozygous aa individuals and assume that the phenotype of AA individuals differs from z_{aa} by an amount s : $z_{AA} = z_{aa} + s$. We write the

phenotype of heterozygous individuals $z_{Aa} = z_{aa} + sh$, h measuring the dominance of A . Defining p_{ij} as the frequency of A in individual ij ($p_{ij} = 0, \frac{1}{2}$, or 1), p_{ij1} and p_{ij2} as the frequencies of A at each of its two homologous genes, and z_{ij} as its phenotype, we have the relation

$$z_{ij} = z_{aa} + s[2hp_{ij} + (1 - 2h)p_{ij1}p_{ij2}]. \quad (11)$$

The expected change in frequency of A over one generation is

$$M_{\delta p} = \frac{1}{2nN} \sum_{i=1}^n \sum_{j=1}^N W_{ij} p_{ij} - p. \quad (12)$$

To the first order in s , this is

$$\begin{aligned} M_{\delta p} &= \frac{s}{2nN} \sum_{i=1}^n \sum_{j=1}^N \frac{dW_{ij}}{ds} p_{ij} + o(s) \\ &= \frac{s}{2nN} \sum_{i=1}^n \sum_{j=1}^N \left(\frac{\partial W_{ij}}{\partial z_{ij}} \frac{dz_{ij}}{ds} + \frac{\partial W_{ij}}{\partial z_i} \frac{dz_i}{ds} + \frac{\partial W_{ij}}{\partial z} \frac{dz}{ds} \right) p_{ij} + o(s), \end{aligned} \quad (13)$$

as W_{ij} depends on z_{ij} , z_i , and z . Equations 11 and 13 give

$$\begin{aligned} M_{\delta p} &= \frac{s}{2} \frac{\partial W_{ij}}{\partial z_{ij}} \left[2h \overline{p_{ij}^2} + (1 - 2h) \overline{p_{ij1} p_{ij2}} \right] \\ &\quad + \frac{s}{2} \frac{\partial W_{ij}}{\partial z_i} \left[2h \overline{p_i^2} + (1 - 2h) \overline{p_{ij1} p_{ij2} p_i} \right] \\ &\quad + \frac{s}{2} \frac{\partial W_{ij}}{\partial z} \left[2h p^2 + (1 - 2h) \overline{p_{ij1} p_{ij2} p} \right] + o(s), \end{aligned} \quad (14)$$

where the overbar means the average among demes and individuals over the whole population.

$M_{\delta p}$ has been defined as the limit of the product $n \times M_{\delta p}$ as n goes to infinity. For this limit to be finite, we have to assume in principle that s is of order $1/n$, so that the product ns has a finite limit as n tends to infinity (for the same reason, s had to be of order $1/N$ in the panmictic case—see previous section). Therefore our method supposes a high number of demes (n high) and weak selection (s small), although simulations will show that these are not major constraints. As s is of order $1/n$, the terms in $o(s)$ of $M_{\delta p}$ vanish when we take the limit of $n \times M_{\delta p}$ as n goes to infinity.

The different averages in Equation 14 can be seen as probabilities of identity: for example, $\overline{p_i^2}$ is the probability of obtaining two A alleles after sampling two genes with replacement from the same deme. In the diffusion limit (n tends to infinity, s tends to zero, ns finite), these probabilities converge to values that depend only on p , the overall frequency of A . Again, we take the example of $\overline{p_i^2}$, the probability that two genes sampled with replacement from the same deme are both A . Consider the two ancestral lineages of these genes backward in time: either these lineages stay and coalesce in the same deme or one of them (or both) migrates to another deme before coalescence occurs; in the first case, the expected coalescence time does not depend on the number of demes n , while in the latter case, coalescence

TABLE 1
Parameters of the model

n	No. of demes in the population
N	No. of adults per deme
m	Migration rate
e	Extinction rate
k	No. of colonizers (per extinct deme)
ϕ	Probability that two genes sampled randomly among the $2k$ gametes recolonizing an extinct deme were in the same deme before migration
ψ_1, ψ_2	Probability that three genes sampled randomly among the $2k$ gametes recolonizing an extinct deme were in the same deme (ψ_1) or in two different demes (ψ_2) before migration
h	Dominance coefficient of allele A over a
s	Phenotypic difference between aa and AA homozygotes
z_{aa}	Phenotype of aa homozygotes

time becomes infinitely long as n tends to infinity. This has been described as a separation of timescales (*e.g.*, WAKELEY 2003 and references therein), because lineages in different demes coalesce at a rate inversely related to n , while genes within demes can coalesce within a few generations, at a rate depending on deme size and migration (roughly, on $m + 1/N$ in the island model), but essentially independent of n . Therefore in the limit as n tends to infinity, at time t independent of n in the past we can consider that the two lineages have either stayed in the same deme and coalesced (with a probability that we call r_1^R) or have migrated to different demes and have not coalesced (with probability $1 - r_1^R$). Thus the probability that our two genes are A is equal to $r_1^R p_t + (1 - r_1^R) p_t^2$, where p_t is frequency of A in the whole population at time t . However, in the limit as n tends to infinity and s tends to zero, the frequency of A in the population does not change over the finite number t of generations, thus $p_t = p$, the present frequency of A . Therefore

$$\begin{aligned} \overline{p_i^2} &= r_1^R p + (1 - r_1^R) p^2 + O(1/n) \\ &= p^2 + r_1^R p q + O(1/n). \end{aligned} \quad (15)$$

This expression, as well as all further expressions involving $O(1/n)$ terms, is an average over all parental populations with the same value of p . These parental populations may differ in the frequency of demes with different copies of the A allele, so that for each possible parental population $\overline{p_i^2}$ may depart from the above average. But the magnitude of these differences vanishes as the number of demes increases, which allows us to use diffusion equations (see WAKELEY 2003, Equation 33, and also ETHIER and NAGYLAKI 1980, Equation 1.5). In the infinite island model, one recovers the classical expression $\overline{p_i^2} = p^2 + r_1^R p q$ for inbreeding coefficients (*e.g.*, CROW and KIMURA 1970). Here r_1^R is equivalent to Wright's F_{ST} (HUDSON 1998; ROUSSET 2002).

The same reasoning is used to calculate the limit of $\overline{p_{ij}^2}$, which is the probability that two genes sampled with replacement in the same individual are both A :

$$\overline{p_{ij}^2} = p^2 + r_0^R p q + O(1/n). \quad (16)$$

Here r_0^R is the probability of coalescence for two genes sampled with replacement from the same individual, in a population with an infinite number of demes.

Denote r_1^D the probability of coalescence of two genes sampled from the same deme *without* replacement, and r_0^D the probability of coalescence of the homologous genes of an individual (still in the limit as n tends to infinity). We have the relations

$$r_1^R = \frac{1}{2N} + \frac{1}{2N} r_0^D + \left(1 - \frac{1}{N}\right) r_1^D \quad (17)$$

and

$$r_0^R = \frac{1}{2} + \frac{1}{2} r_0^D. \quad (18)$$

r_0^D can be shown to be equivalent to Wright's F_{IT} in the infinite island model. As $\overline{p_{j1} p_{j2}}$ is the probability that the two homologous genes of an individual are both A , we have

$$\overline{p_{j1} p_{j2}} = p^2 + r_0^D p q + O(1/n). \quad (19)$$

Last, we need the limit as n tends to infinity of $\overline{p_{j1} p_{j2} p_i}$, which measures the probability that the two homologous genes of an individual and a third gene sampled from the same deme are all A . We call a^R the limit as n tends to infinity of the probability that the three lineages coalesce before one of them moves to another deme, b^R the limit of the probability that only two of them stay in the same deme and coalesce, and $c^R = 1 - a^R - b^R$ the limit of the probability that two lineages move to different demes before any coalescence event has occurred. We have

$$\begin{aligned} \overline{p_{j1} p_{j2} p_i} &= a^R p + b^R p^2 + c^R p^3 + O(1/n) \\ &= p^3 + a^R p q + (1 - c^R) p^2 q + O(1/n). \end{aligned} \quad (20)$$

Denote a^D , b^D , and c^D the same probabilities when the third gene is sampled in a different individual. We have the relations

TABLE 2
Variables of the model

z_{ij}, z_i, z	Phenotype of individual j in deme i , mean phenotype in deme i , and mean phenotype in the population
p_{ij1}, p_{ij2}	Frequency of A at each of the homologous chromosomes of individual j in deme i
p_{ij}, p_i, p	Frequency of A in individual j in deme i , frequency of A deme i , and frequency of A in the population
W_{ij}	Fitness of individual j in deme i
r_0^R, r_0^D	Probability that the ancestral lineages of two genes sampled with (R) or without (D) replacement from the same adult individual coalesce before one of them moves to a different deme, in the limit n goes to infinity
r_1^R, r_1^D	Same probability for two genes sampled with (R) or without (D) replacement from the same deme
a^R, a^D	Probability that the ancestral lineages of the two homologous genes of an individual, plus a third gene sampled with (R) or without (D) replacement from the same deme, all coalesce before one of them moves to a different deme, in the limit as n goes to infinity
b^R, b^D	Probability that only two of the three lineages coalesce before one moves to a different deme, in the same case
c^R, c^D	Probability that two lineages move to different demes before any coalescence has occurred, in the same case

$$a^R = \frac{1}{N}r_0^D + \left(1 - \frac{1}{N}\right)a^D \quad (21)$$

and

$$c^R = \left(1 - \frac{1}{N}\right)c^D. \quad (22)$$

Because adult population size remains constant after regulation, the partial derivatives of the fitness function sum to zero (ROUSSET and BILLIARD 2000); therefore we can express $M_{\delta p}$ as a function of the two first partial derivatives only. After replacing the average products of allele frequencies by the expressions calculated above, we obtain

$$\begin{aligned} M_{\delta p} &= \frac{s}{2} \frac{\partial W_{ij}}{\partial z_{ij}} [h(1 + r_0^D) + (1 - 2h)[r_0^D + (1 - r_0^D)p]]pq \\ &+ \frac{s}{2} \frac{\partial W_{ij}}{\partial z_i} [2hr_1^R + (1 - 2h)[a^R + (1 - c^R - r_0^D)p]]pq + o(1/n). \end{aligned} \quad (23)$$

VARIANCE IN ALLELE FREQUENCY CHANGE

To use the diffusion method, we also need the limit as n tends to infinity of $n \times V_{\delta p}$, where $V_{\delta p} = E[(\delta p)^2 | p]$ and δp is the change in frequency of A over one generation. We have

$$V_{\delta p} = \text{Var}^0[p'] + o(1/n), \quad (24)$$

where $p' = p + \delta p$ and Var^0 is the variance in the neutral case ($s = 0$). We look for an expression for $\text{Var}^0[p']$ to the first order in $1/n$. We have

$$\text{Var}^0[p'] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}^0[p'_i] + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n \text{Cov}^0[p'_i, p'_j]. \quad (25)$$

We show in APPENDIX A that these variance and covariance terms can be expressed, to first order in $1/n$, as functions of the frequency of A in the whole population (p) and an effective population size N_e depending on the various parameters of the model (but not on allele

frequencies). By definition, N_e is the variance effective population size (EWENS 1982).

When $e = 0$ (no extinction), covariances in the change in allele frequencies in different demes equal zero (because sampling occurs independently in each deme). Equations A1–A5 of APPENDIX A lead to

$$V_{\delta p} = \frac{2(1 - r_1^D) - (1 - r_0^D)}{2nN} pq + o(1/n), \quad (26)$$

which is of the form $pq/2N_e$, with

$$N_e = \frac{nN}{2(1 - r_1^D) - (1 - r_0^D)}, \quad (27)$$

a result known under the form $nN/(1 - F_{ST})(1 + F_{IS})$, where $F_{ST} = r_1^D$ and $F_{IS} = (r_0^D - r_1^D)/(1 - r_1^D)$ (see WANG and CABALLERO 1999 for a more general result). In the case of gametic migration, monoecious individuals, and random fertilization, we have $r_0^D = r_1^D$, giving

$$N_e = \frac{nN}{1 - r_1^D}. \quad (28)$$

Extinctions add an additional complication because they introduce correlations in the change in allele frequency in different demes (because the frequency of A in the migrant pool depends on which demes did survive). We show in APPENDIX A that when $e > 0$, the variance effective population size is

$$N_e = \frac{n(1 - e)}{2r_1^R[1 - (1 - e)^2(1 - m)^2]}. \quad (29)$$

This is also the eigenvalue effective size (EWENS 1979, 1982), derived by several independent arguments in ROUSSET (2003).

PROBABILITIES OF COALESCENCE

The expressions for $M_{\delta p}$ and $V_{\delta p}$ derived in the previous sections involve variables representing various probabilities of coalescence in a neutral model, evaluated in the limit as n tends to infinity. These variables can be

calculated by writing recurrence equations (which will depend on the life cycle considered) and calculating values at equilibrium. These recurrence equations are given in APPENDIX B for three different cases: (i) monoecious individuals with random fertilization, gametic migration, and no extinction; (ii) monoecious individuals with rate α of self-fertilization, zygotic migration, and no extinction; and (iii) monoecious individuals with random fertilization, gametic migration, extinction rate e , and gametic recolonization. Here we give only approximate solutions assuming that migration rate is small and deme size large, so that $N = O(1/m)$, and neglecting terms in $O(m)$.

Case i: random fertilization, no extinction: Approximating Equations B5–B9 for small m and large N , one obtains (with $M = Nm$)

$$r_0^D = r_1^D \approx \frac{1}{1 + 4M}, \quad (30)$$

$$(a^D, b^D, c^D) \approx \left(\frac{1}{(1 + 4M)(1 + 2M)}, \frac{6M}{(1 + 4M)(1 + 2M)}, \frac{8M^2}{(1 + 4M)(1 + 2M)} \right). \quad (31)$$

These approximations are also valid for zygotic migration and dioecious individuals (not shown). Although, as found by WHITLOCK (2002), these approximate expressions satisfy the relation $a^D = 2(r_1^D)^2/(1 + r_1^D)$, this is no longer the case when the exact expressions (given in APPENDIX B) are used.

Case ii: partial self-fertilization, no extinction: Here we assume that self-fertilization occurs at a rate α and that fertilization precedes migration. Random fertilization corresponds to $\alpha = 1/N$. We obtain the approximations

$$r_0^D \approx \frac{1 + 2\alpha M}{1 + 2(2 - \alpha)M}, \quad r_1^D \approx \frac{1}{1 + 2(2 - \alpha)M}, \quad (32)$$

$$a^D \approx \frac{1 + \alpha M}{[1 + 2(2 - \alpha)M][1 + (2 - \alpha)M]},$$

$$b^D \approx \frac{2M[3 - \alpha^2 M + 2\alpha(M - 1)]}{[1 + 2(2 - \alpha)M][1 + (2 - \alpha)M]},$$

$$c^D \approx \frac{4(1 - \alpha)(2 - \alpha)M^2}{[1 + 2(2 - \alpha)M][1 + (2 - \alpha)M]}. \quad (33)$$

Case iii: random fertilization, extinctions: We assume here that both migration and recolonization are gametic. In the LIFE CYCLE section, we introduced the parameters k (number of colonizers) and ϕ (probability of common origin). More precisely, ϕ measures the probability that two genes sampled randomly among the $2k$ gametes colonizing an extinct deme were in the same deme before migration. We also need the parameters ψ_1 and ψ_2 , defined as the probabilities that three genes sampled randomly among the $2k$ colonizers were in the same deme before migration (ψ_1) or in two different demes (ψ_2). Following SLATKIN (1977), we distinguish two modes of recolonization: the “migrant pool,” where the $2k$ gametes recolonizing a deme are taken

randomly in the whole population after migration (in this case $\phi = \psi_1 = \psi_2 = 0$), and the “propagule pool,” where the $2k$ gametes were all in the same deme after migration [in this case $\phi = (1 - m)^2$, $\psi_1 = (1 - m)^3$, and $\psi_2 = 3m(1 - m)^2$]. When the extinction rate e is $O(m)$ we obtain

$$r_0^D = r_1^D \approx \frac{1 + eN/k}{1 + 4M + 2eN[1 - \phi(1 - 1/2k)]}. \quad (34)$$

This approximation follows from the recursion (B20) in APPENDIX B and was also obtained by WHITLOCK and McCAULEY (1990). We could not obtain any simple approximation for a^D , b^D , and c^D .

EXAMPLES OF FITNESS FUNCTIONS

We consider two cases of individual selection: “soft” and “hard” selection. The first case corresponds to a regulation of the number of gametes (or zygotes) produced by the different surviving demes just before migration, while in the second case there is no such regulation. However, in both cases there is a phase of population regulation, as N adults are sampled in each deme at the end of every generation. This definition of hard selection is the same as in BARTON (1993), but different from in WHITLOCK (2002), which explains some discrepancies between our results and results in WHITLOCK (2003). In Whitlock’s hard selection model, each deme contributes to the next adult generation in proportion to the number of juveniles produced in the deme before migration; although this definition sounds simple, we do not know any realistic life cycle maintaining a constant population size in this case.

In both cases, we fix z_{aa} to zero and assume that an individual of phenotype z_{ij} produces a number of gametes proportional to $1 + z_{ij}$; the numbers of gametes produced by Aa and AA individuals, relative to aa individuals, are then $1 + sh$ and $1 + s$, respectively.

We note at this point that we need only the limit as n tends to infinity of the fitness function, as terms of order $1/n$ and higher will vanish when we will calculate $\mathcal{M}_{\delta p}$ —see Equation 23.

Soft selection: In this case, we assume that the number of gametes produced in each deme is regulated before migration to a constant value that is the same for all demes; we suppose that this value is still very large. Recall that the fitness of individual ij is the expected number of its gametes that will participate to the next adult generation. We obtain

$$W_{ij} = 2N(1 - e) \left[(1 - m) \frac{1 + z_{ij}}{N(1 + z_i)} + m \frac{1 + z_{ij}}{N(1 + z_i)} + \frac{e}{1 - e} \frac{1 + z_{ij}}{N(1 + z_i)} \right] + O(1/n). \quad (35)$$

Individual ij can participate to the next generation only

if its deme survives (probability $1 - e$). The first term in the parentheses is the frequency of gametes produced by individual ij present in deme i after migration; the second term is the frequency of gametes produced by ij in any other nonextinct deme, after migration, times the number of other nonextinct demes; and the third term is the frequency of gametes produced by ij in the pool of recolonizers, times the number of extinct demes. Simplifying Equation 35 leads to

$$W_{ij} = 2 \frac{1 + z_{ij}}{1 + z_i} + O(1/n), \quad (36)$$

giving $\partial W_{ij}/\partial z_{ij} = 2$, $\partial W_{ij}/\partial z_i = -2$ (derivatives have to be evaluated for $s = 0$, which means here $z_{ij}, z_i, z = 0$).

Hard selection: Here we assume that the number of gametes or juveniles produced in a deme before migration is not regulated: a deme fixed for A produces $1 + s$ times the number of gametes (or juveniles) produced in a deme fixed for a . Here the fitness function is given by

$$W_{ij} = 2N(1 - e) \left[\frac{(1 - m)(1 + z_{ij})}{N[(1 - m)(1 + z_i) + m(1 + z)]} + m \frac{1 + z_{ij}}{N[1 + z]} + \frac{e}{1 - e} \frac{1 + z_{ij}}{N[1 + z]} \right] + O(1/n). \quad (37)$$

Again, the first term in the parentheses represents the proportion of gametes produced by individual ij present in deme i after migration (provided that deme i did not go extinct); the second term is the average proportion of gametes produced by ij in the other nonextinct demes, after migration, times the number of other nonextinct demes; and the third term is the proportion of gametes produced by ij in the pool of recolonizers, times the number of extinct demes (we assume that surviving demes contribute to the pool of colonizers in proportion of the total number of gametes present in the deme). Equation 37 leads to

$$W_{ij} = 2(1 - e) \left[\frac{(1 - m)(1 + z_{ij})}{1 + (1 - m)z_i + mz} + \left(m + \frac{e}{1 - e} \right) \frac{1 + z_{ij}}{1 + z} \right] + O(1/n), \quad (38)$$

giving $\partial W_{ij}/\partial z_{ij} = 2$, $\partial W_{ij}/\partial z_i = -2(1 - e)(1 - m)^2$.

RESULTS

Additivity: In the additive case ($h = 1/2$) and when no extinction occurs ($e = 0$), simple results are obtained. Under hard selection and gametic migration one arrives at

$$G(x) = \exp[-2nNsx], \quad (39)$$

while under soft selection and gametic migration,

$$G(x) = \exp[-2n(N - 1)sx]. \quad (40)$$

The $N - 1$ factor is consistent with the fact that soft

selection means no selection when $N = 1$. In the case of zygotic migration, N has to be replaced by $N(2N)/(2N - 1)$ in Equation 39, while $N - 1$ has to be replaced by $N(2N - 2)/(2N - 1)$ in Equation 40. These differences are significant only for small values of N .

Therefore, in the additive case with no extinction, probabilities of fixation depend neither on population structure nor on inbreeding, as found by MARUYAMA (1970), except for effects of order $1/N$ that have been ignored in previous analyses. Previous simulation studies have also missed them, often because large values of N were assumed. When extinctions occur ($e > 0$), the expression of G is more complex and depends on population structure. For example, under soft selection and haploid migration, we find

$$G(x) = \exp \left\{ -n(N - 1)(1 - e)s \frac{2(1 - r_1^D) - (1 - r_0^D)}{r_1^D N [1 - (1 - e)^2(1 - m)^2]} \right\}. \quad (41)$$

Effects of dominance: When $h \neq 1/2$, the function G is a complicated function of the parameters. Here we present some results relative to case i (monoecious individuals with random fertilization, gametic migration, and no extinction). If we assume that m is small and N large, so that we can use approximations 30 and 31, we obtain

$$G(x) \approx \exp \left\{ -2nNsx \frac{1 + 2Nmx + 4Nmh(1 - x)}{1 + 2Nm} \right\}, \quad (42)$$

which has to be integrated numerically to obtain $u(p)$, the probability of fixation of A . For advantageous mutations ($s > 0$), a crude approximation can be obtained by neglecting terms in x^2 in the expression of G , giving

$$u \left(\frac{1}{2nN} \right) \approx s \frac{1 + 4Nmh}{1 + 2Nm}. \quad (43)$$

We found that this approximation often gives a good idea of the probability of fixation of a rare advantageous allele, except when both the migration rate is large and h is close to zero.

Figure 1 shows the probability of fixation of a rare advantageous allele as a function of its dominance coefficient, for different values of the migration rate. Here we used the exact expressions of the probabilities of identity for case i (given in APPENDIX B); however, Equation 43 gives very close results (not shown). We used the soft selection model, but under hard selection the results are qualitatively and quantitatively very similar for large N . Figure 1 illustrates the fact that population structure increases the probability of fixation of recessive advantageous mutants, while it decreases the probability of fixation of dominant advantageous mutants. It also shows that the diffusion approximation gives poor results for small values of m ; in the last case SLATKIN's (1981) low-

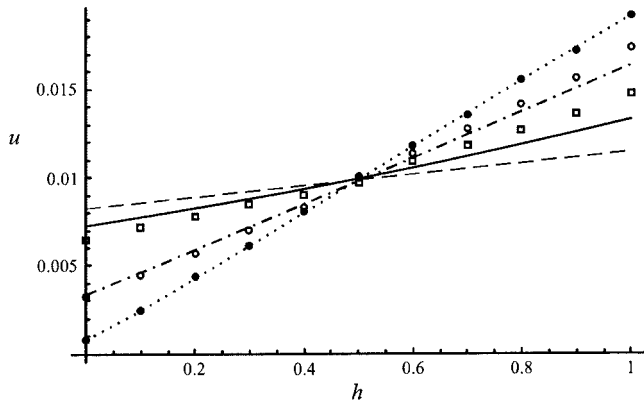


FIGURE 1.—Probability of fixation of an advantageous mutation ($s = 0.01$) when present as a single copy, as a function of its dominance coefficient (h). Migration is gametic, individuals are monoecious with random fertilization, and no extinction occurs ($e = 0$). Number of demes, $n = 200$; deme size, $N = 100$. Results from the diffusion model are: dotted line, migration rate $m = 0.1$; dotted/dashed line, $m = 0.01$; dashed line, $m = 0.001$. Simulation results are: solid circles, $m = 0.1$; open circles, $m = 0.01$; squares, $m = 0.001$. The solid line corresponds to SLATKIN's (1981) low migration limit.

migration approximation performs better (this is discussed later).

Effects of selfing: In case ii (monoecious individuals with rate α of self-fertilization, zygotic migration, and no extinction), approximations (32) and (33) lead to

$$G(x) = \exp\left[-2nNsx\left[2h + (1-2h)\frac{1 + (2-\alpha)Nm[\alpha + (1-\alpha)x]}{1 + (2-\alpha)Nm}\right]\right], \quad (44)$$

which has to be integrated numerically. Neglecting the terms in x^2 , we obtain a crude approximation for advantageous mutations:

$$u\left(\frac{1}{2nN}\right) \approx s \frac{1 + (2-\alpha)[\alpha + 2h(1-\alpha)]Nm}{1 + (2-\alpha)Nm}. \quad (45)$$

Figure 2 shows the effect of partial self-fertilization on the probability of fixation of a recessive advantageous mutation ($h = 0$). Again, we used the exact expressions of the probabilities of identity and integrated numerically the G function, but Equation 45 gives very close results. We still observe a discrepancy between analytical and simulation results at small values of m .

Effects of the extinction rate: Figure 3 shows some results concerning case iii (monoecious individuals with random fertilization, gametic migration, extinction rate e , and gametic recolonization), taking the example of an advantageous underdominant mutant. Underdominance corresponds to the case where heterozygotes are disadvantaged ($h < 0$); it can result from chromosomal rearrangements (LANDE 1985) and possibly from mutations affecting transcriptional regulation during the development (GIBSON 1996). In a large panmictic population, such mutations have a very low probability of fixation

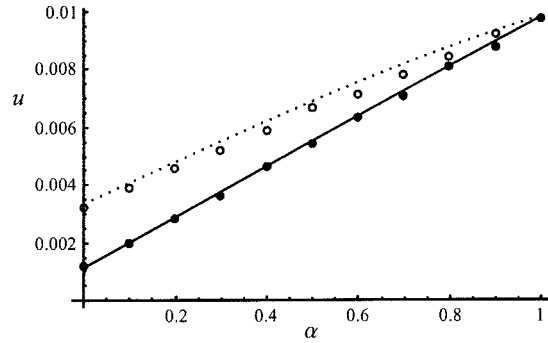


FIGURE 2.—Probability of fixation of a recessive advantageous mutation ($s = 0.01$, $h = 0$), as a function of the rate of self-fertilization (α). Individuals are monoecious and migration is zygotic. Number of demes, $n = 200$; deme size, $N = 100$; no extinction occurs ($e = 0$). Results from the diffusion model are: dotted line, migration rate $m = 0.01$; solid line, migration rate $m = 0.05$. Simulation results are: open circles, $m = 0.01$; solid circles, $m = 0.05$.

(e.g., KIMURA 1962). Using a low-migration limit, LANDE (1985) showed that local extinctions can increase the probability of fixation of underdominant mutations, in an island model of population structure. This effect has also been observed by simulation in a two-dimensional stepping-stone model (MICHALAKIS and OLIVIERI 1993).

Figure 3 represents the probability of fixation of an underdominant mutation, relative to a neutral mutation, as a function of the extinction rate e , and for different values of k , the number of colonizers. Figure 3A corresponds to the migrant pool model of colonization, while Figure 3B corresponds to the propagule pool model. We used the soft selection model, but again results are very similar with hard selection (not shown). The effect of the rate of extinction on the fixation probability of the mutant is nonmonotonic. This may result from two opposite forces: extinctions increase the probability that an advantageous AA genotype will appear before allele A is lost, in particular when the number of colonizers is small, but also decrease the chance of fixation of A once AA genotypes have been created (by increasing random drift, extinctions reduce the probability of fixation of advantageous alleles).

DISCUSSION

Figures 1 and 2 show that the analytical model gives good predictions when the migration rate is not too small, while for small values of m predictions are not as good (except when $h = \frac{1}{2}$ and $e = 0$, since in that case we obtain an expression independent on m for the probability of fixation). This comes from the fact that we calculated the probabilities of coalescence before migration (r_0 , r_1 , a , b , c) in the neutral case ($s = 0$); as s increases, these probabilities become more and more different from the neutral values, and this difference depends on the migration rate. As previously noted,

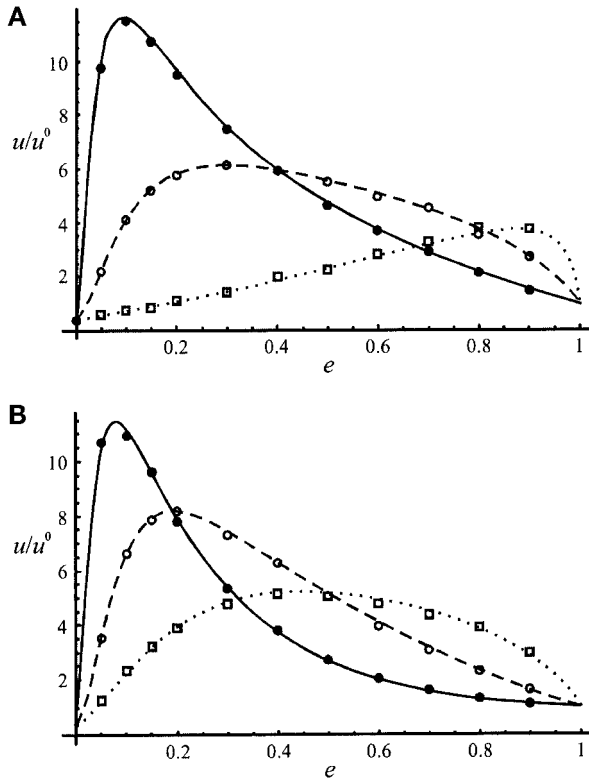


FIGURE 3.—Probability of fixation of an underdominant advantageous mutation ($s = 0.02$, $h = -0.1$) relative to a neutral mutation, as a function of the extinction rate (e). Individuals are monoecious with random fertilization, and migration and recolonization are gametic; number of demes, $n = 200$; deme size, $N = 100$; migration rate, $m = 0.1$. A corresponds to the “migrant pool” model and B to the “propagule pool” model. Results from the diffusion model are: solid line, number of colonizers $k = 2$; dashed line, $k = 10$; dotted line, $k = 100$. Simulation results are: solid circles, $k = 2$; open circles, $k = 10$; squares, $k = 100$.

what matters in the argument based on a separation of timescales is that changes in allele frequencies within demes, due to drift and migration, are fast relative to selection (see also WAKELEY 2003), hence that s is small relative to $m + 1/N$. In practice, the validity of the diffusion approximation depends on weak selection (small s) and not-too-small migration (roughly, $m > s$). This can also be expressed in terms of our coalescent argument: for relatively high values of m , two lineages from the same deme either separate to different demes (in the past) or coalesce rapidly within the deme, and during this short time selection does not have much effect. For small values of the migration rate, however, the mean within-deme coalescence time becomes longer, and the effect of selection during that time cannot be neglected any more.

Although the validity of the diffusion approximation depends on a large number of demes (large n) and weak selection (small s), these two conditions are in fact not very restrictive: Table 3 shows that our solution

TABLE 3

Accuracy of the diffusion results

	$s = 0.2$	$s = 0.1$	$s = 0.05$	$s = 0.01$
$n = 5$	0.020953 0.018484	0.013665 0.012514	0.009097 0.008583	0.003737 0.003642
$n = 10$	0.016542 0.014920	0.010498 0.009834	0.006839 0.006472	0.002721 0.002705
$n = 20$	0.013498 0.012631	0.008288 0.007933	0.005254 0.005151	0.002003 0.001974
$n = 50$	0.010930 0.010594	0.006383 0.006179	0.003870 0.003783	0.001369 0.001343

Probability of fixation of allele A when present as a single copy, for different values of n and s . Here $e = 0$ (no extinction), $N = 100$, $h = 0$, and $m = 0.1$; individuals are monoecious, fertilization is random and migration is gametic. The value at the top of each cell is the diffusion result, and the value at the bottom has been obtained by simulation (5 million replicates for each set of parameters).

gives accurate results for values of n as low as 10 and values of s as high as 0.1. When the exact expressions of the probabilities of identity are used, no other assumption needs to be made; for example, deme size (N) and rate of extinction (e) can be arbitrarily large or small. This is a difference between the approach used in CHERRY and WAKELEY (2003), CHERRY (2003a), and WHITLOCK (2003) and our approach. Indeed, these authors calculate various quantities that are similar to the probabilities of identity in Equation 14 by using the fact that the distribution of allele frequencies among demes can be approximated by a β -distribution when N is large and m is small. From our coalescent argument, we obtain the first moments of the distribution of allele frequencies among demes without having to assume that N is large or m small. In fact, CHERRY (2003a) and WHITLOCK's (2003) approximations for the G function are the same as our Equation 42, which assumes small m and large N . We found that Equation 42 gives satisfying results as long as deme size is not too small; for small values of N , however, using the exact expressions of r_0 , r_1 , a , and c gives more precise results. For example, with $n = 50$, $N = 15$, and $m = 0.3$, Equation 42 predicts that the probability of fixation of a deleterious allele with $s = -0.005$ and $h = 0$ (relative to a neutral allele) is 0.0072, while using the exact expressions of r_1 , a , and c gives a value of 0.01128, and simulations give 0.01121 (for soft selection). Although assuming either hard or soft selection does not make much difference when N is large and m small (indeed, Equation 42 holds in both cases), this is not the case when N is small, as can be shown by using expressions of the G function that do not make any assumption on N and m —obtained from equations B5–B9 and 36 and 38. This is illustrated by Figure 4, which shows the probability of fixation of a recessive advantageous allele obtained from these expressions

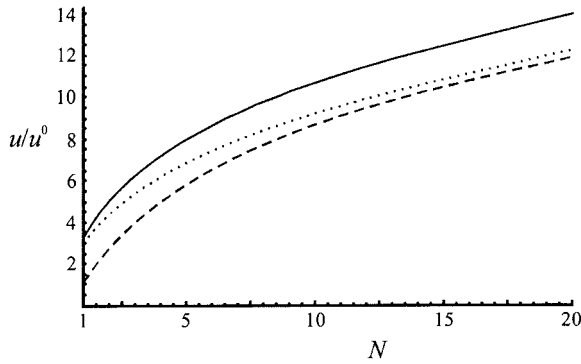


FIGURE 4.—Probability of fixation of an advantageous recessive mutation ($h = 0$, $s = 0.02$), relative to a neutral mutation, as a function of deme size (N). Migration is gametic, individuals are monoecious with random fertilization, and no extinction occurs ($e = 0$). Number of demes, $n = 100$; migration rate, $m = 0.3$. Dashed line, soft selection; dotted line, hard selection; solid line, approximation (42) for N large and m small.

and from Equation 42, for low values of N . These examples show that using exact values of the probabilities of coalescence leads to more precise approximations; moreover, exact values of r_0 , r_1 , a , and c can be calculated without too much difficulty for many different life cycles, while it is not always easy to obtain approximations for the distribution of allele frequencies among demes (for example, when extinctions occur).

BARTON (1993) obtained approximations for the probability of fixation of a single mutant in a model with extinction and additive selection, assuming that deme size is large. We observed that his approximation for the case $\phi = 0$ (Equation 11b in BARTON 1993) gives similar results as our solution when migration and extinction rates are small (not shown); however, in our model recolonization does not occur exactly as in BARTON (1993); therefore, one must be cautious when comparing our results.

An important consequence of population subdivision is an increase of the fixation probability of recessive advantageous mutations. Figure 5 shows this effect for a population of size 10,000; the different curves correspond to different modes of subdivision (10 demes of 1000 individuals, 100 demes of 100 individuals, and 1000 demes of 10 individuals). One can see that the increase in fixation probability of the recessive mutation is substantial only with “strong” population structure (small migration rate and/or small deme size); a moderate structure such as $N = 100$, $m = 0.1$ has virtually no effect on the fixation probability of recessive alleles.

We have demonstrated a simple method to construct diffusion approximations for populations subdivided according to an island model. It is based on the separation of timescales and on the additional fact that for the different life cycles, expressions for $M_{\delta p}$ can be deduced

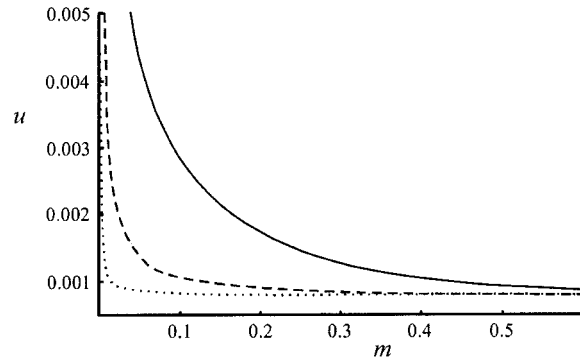


FIGURE 5.—Effect of population structure on the probability of fixation of a recessive advantageous mutation ($s = 0.01$, $h = 0$). x -axis, migration rate; y -axis, probability of fixation of the mutation when present as a single copy. Migration is gametic and fertilization random; extinctions do not occur ($e = 0$). Solid line, $n = 1000$, $N = 10$; dashed line, $n = 100$, $N = 100$; dotted line, $n = 10$, $N = 1000$.

mechanically from the fitness function with minimal risk of error by the “direct fitness” method, where the fitness function itself is a direct expression of the life cycle considered. This method has been useful for analyzing selection on various traits under different population structures (*e.g.*, TAYLOR and FRANK 1996; ROUSSET and GANDON 2002). In this context, fitness has to be defined as the expected number of “successful” gametes of an individual, *i.e.*, those that effectively participate to the gene pool of the next adult generation. Many different uses of the word “fitness” can be found in the literature; it often means “relative fecundity” or “relative survival.” Defining a fitness function as we do proves to be useful since it leads to general expressions like Equation 23.

The direct fitness method used here has also proven convenient to study kin selection (TAYLOR and FRANK 1996; ROUSSET and BILLIARD 2000). A substantial part of kin selection theory has been devoted to the evolution of altruistic traits; in some cases, it can be shown that such traits always increase in frequency if a condition of the form $-c + rb > 0$ holds, where c and b are the cost and benefit of the altruistic act, and r the “relatedness” between interactants, assumed to be independent of the frequency of the allele coding for altruism (HAMILTON 1964). Whether such a condition exists is not always obvious, and the present application of the direct fitness formalism may clarify such problems, since from Equation 23 one easily obtains conditions under which an allele increases in frequency in a structured population. For example, when $h = 1/2$, Equation 23 shows that $M_{\delta p}$ is always positive provided that

$$\frac{\partial W_{ij}}{\partial z_{ij}} + \frac{2r_1^R}{1 + r_0^D} \frac{\partial W_{ij}}{\partial z_i} > 0. \quad (46)$$

This is the correct form of Hamilton’s rule that one

obtains by modern inclusive fitness techniques; it shows that $-c$, b , and r have to be defined as $\partial W_{ij}/\partial z_{ij}$, $\partial W_{ij}/\partial z_i$, and $2r_1^R/(1 + r_0^D)$, respectively. When $h \neq 1/2$ we can also define a relatedness coefficient from Equation 23, but now this coefficient depends on p , the frequency of A . Conversely, our model can be used to include the effects of genetic drift in kin selection models and derive fixation probabilities and fixation times of alleles coding for altruistic behaviors.

Many other applications are outside the scope of this article, but are worth mentioning. For example, one can compute the distribution of allele frequency of recessive deleterious alleles in a metapopulation and consider the implication of different demographic regimes on the purge of such alleles. Average times to fixation or loss of new alleles are easily obtained once the expressions of $M_{\delta p}$ and $V_{\delta p}$ are known (CROW and KIMURA 1970); although in this article we focused on fixation probabilities, fixation times are equally important in determining the rate of evolution of populations. CHERRY and WAKELEY (2003), CHERRY (2003b), and WHITLOCK (2003) showed that the diffusion method provides good approximations for fixation times in structured populations. In general, a lower rate of migration slows down fixation, while extinction/recolonization can speed it up.

One may wonder about the generality of the present approach with respect to the type of population structure. Although in this article we stayed in the framework of the island model, the methodology presented here can be used with other spatial models, as long as coalescence occurs at two different timescales: a fast one during which selection can be neglected and a slow one that becomes infinitely long as population size tends to infinity. This is the case, for example, for cytoplasmic genes: coalescence is fast if lineages are sampled within the same organism and slow if they are sampled in different organisms. The two scales can also be family *vs.* population (*e.g.*, CABALLERO and HILL 1992). In these examples many questions arise about conflicting levels of selection; multilevel selection is easily incorporated into this framework by redefining the fitness function.

We thank Joshua Cherry, John Wakeley, and Michael Whitlock for providing articles in press and Joshua Cherry, Montgomery Slatkin, John Wakeley, Michael Whitlock, and an anonymous reviewer for helpful comments on the manuscript. This is paper ISEM 2003–79.

LITERATURE CITED

- BARTON, N. H., 1993 The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.* **62**: 149–157.
- CABALLERO, A., and W. G. HILL, 1992 Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* **131**: 493–507.
- CARR, R. N., and R. F. NASSAR, 1970 Effects of selection and drift on the dynamics of finite populations. I. Ultimate probability of fixation of a favorable allele. *Biometrics* **26**: 41–49.
- CHERRY, J. L., 2003a Selection in a subdivided population with dominance or local frequency dependence. *Genetics* **163**: 1511–1518.
- CHERRY, J. L., 2003b Selection in a subdivided population with local extinction and recolonization. *Genetics* **164**: 789–795.
- CHERRY, J. L., and J. WAKELEY, 2003 A diffusion approximation for selection and drift in a subdivided population. *Genetics* **163**: 421–428.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- ETHIER, S. N., and T. NAGYLAKI, 1980 Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.* **12**: 14–49.
- EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- EWENS, W. J., 1982 On the concept of effective population size. *Theor. Popul. Biol.* **21**: 373–378.
- GALE, J. S., 1990 *Theoretical Population Genetics*. Unwin Hyman, London.
- GIBSON, G., 1996 Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor. Popul. Biol.* **49**: 58–89.
- HAMILTON, W. D., 1964 The genetical evolution of social behaviour. *J. Theor. Biol.* **7**: 1–52.
- HUDSON, R. R., 1998 Island models and the coalescent process. *Mol. Ecol.* **7**: 413–418.
- KARLIN, S., and H. M. TAYLOR, 1981 *A Second Course in Stochastic Processes*. Academic Press, New York.
- KIMURA, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719.
- KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Prob.* **1**: 177–232.
- LANDE, R., 1985 The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* **54**: 323–332.
- MARUYAMA, T., 1970 On the fixation probability of mutant genes in a subdivided population. *Genet. Res.* **15**: 221–225.
- MICHALAKIS, Y., and I. OLIVIERI, 1993 The influence of local extinctions on the probability of fixation of chromosomal rearrangements. *J. Evol. Biol.* **6**: 153–170.
- PANNELL, J. R., and B. CHARLESWORTH, 1999 Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664–676.
- ROUSSET, F., 2002 Inbreeding and relatedness coefficients: what do they measure? *Heredity* **88**: 371–380.
- ROUSSET, F., 2003 Effective size in simple metapopulation models. *Heredity* **91**: 107–111.
- ROUSSET, F., 2004 *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
- ROUSSET, F., and S. BILLIARD, 2000 A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**: 814–825.
- ROUSSET, F., and S. GANDON, 2002 Evolution of the distribution of dispersal distance under distance-dependent cost of dispersal. *J. Evol. Biol.* **15**: 515–523.
- SLATKIN, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Popul. Biol.* **12**: 253–262.
- SLATKIN, M., 1981 Fixation probabilities and fixation times in a subdivided population. *Evolution* **35**: 477–488.
- TAKAHATA, N., 1991 Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. *Genetics* **129**: 585–595.
- TAYLOR, P. D., and S. A. FRANK, 1996 How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.
- WADE, M. J., and D. E. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**: 995–1005.
- WAKELEY, J., 2003 Polymorphism and divergence for island-model species. *Genetics* **163**: 411–420.
- WANG, J., and A. CABALLERO, 1999 Developments in predicting the effective population size of subdivided populations. *Heredity* **82**: 212–226.
- WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191–1202.
- WHITLOCK, M. C., 2003 Fixation probability and time in a metapopulation. *Genetics* **164**: 767–779.
- WHITLOCK, M. C., and N. H. BARTON, 1997 The effective size of a subdivided population. *Genetics* **146**: 427–441.

WHITLOCK, M. C., and D. E. McCAULEY, 1990 Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717–1724.

Communicating editor: L. EXCOFFIER

APPENDIX A

No extinction: When $e = 0$, $V_{\delta p}$ is given by

$$V_{\delta p} = \frac{1}{n^2} \sum_{i=1}^n \text{Var}^0[p'_i] + o(1/n) \quad (\text{A1})$$

since the covariance in the change in allele frequency in two different demes equals zero (sampling occurs independently in each deme). We have

$$\text{Var}^0[p'_i] = E^0[p'_i]E^0[q'_i] - E^0[p'_i q'_i] \quad (\text{A2})$$

and

$$E^0[p'_i] = (1 - m)p_i + mp \quad (\text{A3})$$

and since the averages over i of $E^0[p'_i q'_i]$ and $p'_i q'_i$ both tend to $(1 - r_1^R)pq$ as n tends to infinity, we obtain

$$\begin{aligned} V_{\delta p} &= \frac{1}{n} [1 - (1 - m)^2] r_1^R pq + o(1/n) \\ &= (r_1^R - r_1^D) \frac{pq}{n} + o(1/n) \end{aligned} \quad (\text{A4})$$

as $r_1^D = (1 - m)^2 r_1^R$. Using Equation 17, one arrives at

$$V_{\delta p} = \frac{2(1 - r_1^D) - (1 - r_1^D)}{2nN} pq + o(1/n) \quad (\text{A5})$$

and thus we can define a variance effective population size as

$$N_e = \frac{nN}{2(1 - r_1^D) - (1 - r_1^D)} \quad (\text{A6})$$

so that

$$V_{\delta p} = \frac{pq}{2N_e} + o(1/n). \quad (\text{A7})$$

Extinction-recolonization: Extinctions introduce correlations in the change in allele frequency in different demes (because the frequency of A in the migrant pool depends on which demes did survive). $V_{\delta p}$ is expressed as

$$V_{\delta p} = \frac{1}{n^2} \sum_{i=1}^n \text{Var}^0[p'_i] + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \text{Cov}^0[p'_i, p'_j] + o(1/n). \quad (\text{A8})$$

The first term is obtained as above. $\text{Var}^0[p'_i]$ is still given by Equation A2, with

$$E^0[p'_i] = (1 - e)[(1 - m)p_i + mp] + ep, \quad (\text{A9})$$

giving

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var}^0[p'_i] = \frac{1}{n} [1 - (1 - e)^2(1 - m)^2] r_1^R pq + o(1/n). \quad (\text{A10})$$

We then need to calculate the average covariance of the change in frequency of A in different demes, to the first order in $1/n$. We start by noting that at the beginning of the life cycle, the probability of sampling two A alleles with replacement is p^2 if we sample them from the whole population, $\overline{p_i^2}$ if we sample them from the same deme, and $\overline{p_i p_j}$ if we sample them from different demes ($i \neq j$). This gives

$$p^2 = \frac{1}{n} \overline{p_i^2} + \left(1 - \frac{1}{n}\right) \overline{p_i p_j} \quad (\text{A11})$$

and since $\overline{p_i^2}$ tends to $p^2 + r_1^R pq$ as n tends to infinity,

$$\overline{p_i p_j} = p^2 - \frac{1}{n} r_1^R pq + o(1/n). \quad (\text{A12})$$

We have

$$\text{Cov}[p'_i, p'_j] = E[p'_i p'_j] - E[p'_i]E[p'_j]. \quad (\text{A13})$$

From Equations A9 and A12 we obtain the second term of the covariance:

$$\frac{1}{n^2} \sum_{i \neq j} E[p'_i]E[p'_j] = p^2 - \frac{1}{n} (1 - e)^2 (1 - m)^2 r_1^R pq + o(1/n). \quad (\text{A14})$$

The first term of the covariance is the probability of sampling two A alleles from two different demes, at the next generation. With a probability of order $1/n$, these two lineages come from the same deme in the previous generation and are both A with probability $p^2 + r_1^R pq + O(1/n)$, while with a probability of order 1 they come from different demes, in which case the probability that they are A is given by Equation A12. We calculate the probability that they come from the same deme as follows: if the two demes where the genes are sampled have not become extinct in the previous generation [probability $(1 - e)^2$], the probability that the two lineages come from a single deme in the previous generation is

$$\frac{1 - (1 - m)^2}{n(1 - e)} + o(1/n), \quad (\text{A15})$$

while if one or both demes became extinct [probability $1 - (1 - e)^2$], this probability is $1/[n(1 - e)]$. Taking all cases into account, one arrives at

$$\frac{1}{n^2} \sum_{i \neq j} E[p'_i p'_j] = p^2 + \left[\frac{1 - (1 - e)^2(1 - m)^2}{n(1 - e)} \frac{1}{n} \right] r_1^R pq + o(1/n), \quad (\text{A16})$$

which, combined with Equations A10 and A14, finally gives

$$V_{\delta p} = \frac{1 - (1 - e)^2(1 - m)^2}{n(1 - e)} r_1^R pq + o(1/n). \quad (\text{A17})$$

Therefore, when extinctions occur, the variance effective population size is

$$N_e = \frac{n(1-e)}{2r_1^R[1-(1-e)^2(1-m)^2]}. \quad (\text{A18})$$

APPENDIX B

Case i: monoecious individuals with random fertilization, gametic migration, and no extinction: The recurrence equations are

$$r_0^{D'} = r_1^{D'} = (1-m)^2 \left(\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) r_1^D \right) \quad (\text{B1})$$

$$a^{D'} = (1-m)^3 \left[\frac{1}{(2N)^2} + \frac{3}{2N} \left(1 - \frac{1}{2N}\right) r_1^D + \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) a^D \right] \quad (\text{B2})$$

$$b^{D'} = 3m(1-m)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) r_1^D \right] + (1-m)^3 \left[\frac{3}{2N} \left(1 - \frac{1}{2N}\right) (1 - r_1^D) + \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) b^D \right] \quad (\text{B3})$$

$$c^{D'} = m^3 + 3m^2(1-m) + 3m(1-m)^2 \left(1 - \frac{1}{2N}\right) (1 - r_1^D) + (1-m)^3 \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) c^D, \quad (\text{B4})$$

which gives

$$r_0^D = r_1^D = \frac{(1-m)^2}{2N - (1-m)^2(2N-1)} \quad (\text{B5})$$

$$a^D = \frac{(1-m)^3}{Y} [N + (2N-1)(1-m)^2] \quad (\text{B6})$$

$$b^D = \frac{3m(1-m)^2 N}{Y} [2N + (1-m)(2-m)(2N-1)] \quad (\text{B7})$$

$$c^D = \frac{2m^2 N^2}{Y} [2N(3-2m) + (3-m)(1-m)^2(2N-1)] \quad (\text{B8})$$

with

$$Y = [2N - (2N-1)(1-m)^2][2N^2 - (2N-1)(N-1)(1-m)^3]. \quad (\text{B9})$$

Case ii: monoecious individuals with rate α of self-fertilization, zygotic migration, and no extinction: Because we assume zygotic migration, we need to calculate the probabilities that three genes sampled in three different individuals from the same deme all have a common ancestor (d^D), that two of them have a common ancestor (e^D), and that they have no common ancestor (f^D), in the infinite island model. The recurrence equations are

$$r_0^{D'} = (1-\alpha)r_1^D + \alpha \left(\frac{1}{2} + \frac{1}{2} r_0^D \right) \quad (\text{B10})$$

$$r_1^{D'} = (1-m)^2 \left[\frac{1}{2N} + \frac{1}{2N} r_0^D + \left(1 - \frac{1}{N}\right) r_1^D \right] \quad (\text{B11})$$

$$a^{D'} = (1-m)^2(1-\alpha) \left(\frac{1}{N} r_1^D + \frac{1}{N} a^D + \left(1 - \frac{2}{N}\right) a^D \right) + (1-m)^2 \alpha \left(\frac{1}{4N} + \frac{3}{4N} r_0^D + \frac{1}{2} \left(1 - \frac{1}{N}\right) (r_1^D + a^D) \right) \quad (\text{B12})$$

$$b^{D'} = [1 - (1-m)^2] \left[(1-\alpha)r_1^D + \alpha \left(\frac{1}{2} + \frac{1}{2} r_0^D \right) \right] + (1-m)^2(1-\alpha) \left[\frac{1}{N} (1 - r_1^D) + \frac{1}{N} b^D + \left(1 - \frac{2}{N}\right) e^D \right] + (1-m)^2 \alpha \left[\frac{3}{4N} (1 - r_0^D) + \frac{1}{2} \left(1 - \frac{1}{N}\right) (1 - r_1^D + b^D) \right] \quad (\text{B13})$$

$$c^{D'} = [1 - (1-m)^2] \left[(1-\alpha)(1 - r_1^D) + \alpha \frac{1}{2} (1 - r_0^D) \right] + (1-m)^2 \left[(1-\alpha) \left(\frac{1}{N} c^D + \left(1 - \frac{2}{N}\right) f^D \right) + \alpha \frac{1}{2} \left(1 - \frac{1}{N}\right) c^D \right] \quad (\text{B14})$$

$$d^{D'} = (1-m)^3 \left[\frac{1}{(2N)^2} + \frac{3}{(2N)^2} r_0^D + \frac{3}{2N} \left(1 - \frac{1}{N}\right) (r_1^D + a^D) + \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) d^D \right] \quad (\text{B15})$$

$$e^{D'} = 3m(1-m)^2 \left[\frac{1}{2N} + \frac{1}{2N} r_0^D + \left(1 - \frac{1}{N}\right) r_1^D \right] + (1-m)^3 \left[\frac{3}{(2N)^2} (1 - r_0^D) + \frac{3}{2N} \left(1 - \frac{1}{N}\right) (1 - r_1^D + b^D) + \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) e^D \right] \quad (\text{B16})$$

$$f^{D'} = m^3 + 3m^2(1-m) + 3m(1-m)^2 \left[\frac{1}{2N} (1 - r_0^D) + \left(1 - \frac{1}{N}\right) (1 - r_1^D) \right] + (1-m)^3 \left[\frac{3}{2N} \left(1 - \frac{1}{N}\right) c^D + \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) f^D \right], \quad (\text{B17})$$

which gives for r_0^D and r_1^D

$$r_0^D = \frac{1 - [1 - (1-m)^2][1 - \alpha N]}{1 - [1 - (1-m)^2][1 - (2-\alpha)N]}, \quad (\text{B18})$$

$$r_1^D = \frac{(1-m)^2}{1 - [1 - (1-m)^2][1 - (2-\alpha)N]}. \quad (\text{B19})$$

The exact expressions of a^D , b^D , and c^D are complicated and not given here for space reasons, but are available by request to the authors.

Case iii: monoecious individuals with random fertilization, gametic migration, extinction rate e , and gametic recolonization: When extinctions occur, the recurrence equations become

$$r_0^{D'} = r_1^{D'} = \left[(1-e)(1-m)^2 + e \left(1 - \frac{1}{2k}\right) \phi \right] \times \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) r_1^D \right] + \frac{e}{2k} \quad (\text{B20})$$

(see ROUSSET 2003).

$$\begin{aligned}
 a^{\text{D}'} &= \left[(1-e)(1-m)^3 + e \left(1 - \frac{1}{2k} \right) \left(1 - \frac{1}{k} \right) \psi_1 \right] \\
 &\times \left[\frac{1}{(2N)^2} + \frac{3}{2N} \right] \left(1 - \frac{1}{2N} \right) r_1^{\text{D}} + \left(1 - \frac{1}{2N} \right) \left(1 - \frac{1}{N} \right) a^{\text{D}} \right] \\
 &+ e \left[\frac{1}{(2k)^2} + \frac{3}{2k} \right] \left(1 - \frac{1}{2k} \right) \phi \left(\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) r_1^{\text{D}} \right) \right] \quad (\text{B21})
 \end{aligned}$$

$$\begin{aligned}
 b^{\text{D}'} &= \left[(1-e)(1-m)^3 + e \left(1 - \frac{1}{2k} \right) \left(1 - \frac{1}{k} \right) \psi_1 \right] \\
 &\times \left[\frac{3}{2N} \left(1 - \frac{1}{2N} \right) (1 - r_1^{\text{D}}) + \left(1 - \frac{1}{2N} \right) \left(1 - \frac{1}{N} \right) b^{\text{D}} \right] \\
 &+ \left[3(1-e)m(1-m)^2 + e \left(1 - \frac{1}{2k} \right) \left(1 - \frac{1}{k} \right) \psi_2 \right] \left[\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) r_1^{\text{D}} \right] \\
 &+ e \frac{3}{2k} \left(1 - \frac{1}{2k} \right) \left[(1-\phi) + \phi \left(1 - \frac{1}{2N} \right) (1 - r_1^{\text{D}}) \right] \quad (\text{B22})
 \end{aligned}$$

$$\begin{aligned}
 c^{\text{D}'} &= (1-e) \left[\begin{aligned} &m^3 + 3m^2(1-m) + 3m(1-m)^2 \left(1 - \frac{1}{2N} \right) (1 - r_1^{\text{D}}) \\ &+ (1-m)^3 \left(1 - \frac{1}{2N} \right) \left(1 - \frac{1}{N} \right) c^{\text{D}} \end{aligned} \right] \\
 &+ e \left(1 - \frac{1}{2k} \right) \left(1 - \frac{1}{k} \right) \left[\begin{aligned} &1 - \psi_1 - \psi_2 + \psi_2 \left(1 - \frac{1}{2N} \right) (1 - r_1^{\text{D}}) \\ &+ \psi_1 \left(1 - \frac{1}{2N} \right) \left(1 - \frac{1}{N} \right) c^{\text{D}} \end{aligned} \right] \quad (\text{B23})
 \end{aligned}$$

Here again, we give only the solution for $r_0^{\text{D}} = r_1^{\text{D}}$, while the solutions for a^{D} , b^{D} , and c^{D} are available on request:

$$r_0^{\text{D}} = r_1^{\text{D}} = \frac{2k(1-m)^2(1-e) + 2Ne + (2k-1)e\phi}{2k[2N - (2N-1)(1-e)(1-m)^2] - (2k-1)(2N-1)e\phi}. \quad (\text{B24})$$